Identifying a Protein by MALDI–TOF Mass Spectrometry W

An Experiment for the Undergraduate Laboratory

Anne E. Counterman,* Matthew S. Thompson, and David E. Clemmer
Department of Chemistry, Indiana University, Bloomington, IN 47405; *aec14@psu.edu

Mass spectrometry has become a widely used tool for protein sequencing and fingerprinting *(1)*. This is largely due to the introduction of soft ionization sources for analysis of biomolecules, such as electrospray ionization *(2)* and matrix-assisted laser desorption ionization (MALDI, ref *3*). The combination of these sources with time-of-flight (TOF) mass spectrometry makes this technology affordable for many research laboratories and university mass spectrometry facilities. For those not familiar with the technique, we provide a brief overview of the operating principles in the experimental section. Additionally, an excellent tutorial on MALDI–TOF mass spectrometry has previously been given in this *Journal (4)*.

One standard methodology for protein identification using mass spectrometry involves digesting a protein with an enzyme such as trypsin, chymotrypsin, or endoproteinases Asp-N or Glu-C (each of which cleave proteins at specific residue types) to form a mixture of peptides. Trypsin, the most frequently used proteolytic enzyme, cleaves proteins on the C-terminal side of lysine (K) and arginine (R) residues, as illustrated below (lines denote cleavage sites; letters are the abbreviation for the amino acids)

AFDEK⎦AWSVAR⎦VVAGVANALAHK⎦EK⎦FVIEIR⎦....

Peptides generated in this manner have a set of masses that is characteristic of the protein from which they are derived—a fingerprint. The mixture of peptides produced upon digestion can be analyzed rapidly (in minutes) by MALDI–TOF. Masses of the peptides can be compared against databases of digest masses from all proteins with known sequences (and, in some cases, putative proteins translated from DNA sequences, ref *5, 6*). The database searches yield a list of proteins that are most consistent with the data, as well as a statistical ranking of the matches. In order to assess the degree of confidence associated with a protein identification (and determine the protein rankings), it is important to consider not only the number of masses that can be assigned to expected digestion products, but also the sequence coverage achieved (i.e., the fraction of residues assigned) and the frequency with which each given mass occurs in the database.

This experiment requires students to use mass spectral data to find the identity of an unknown protein sample by performing a database search on the Internet. Students learn the fundamental operating principles of MALDI–TOF, are exposed to a commonly used technique for protein identification, and use the Internet to perform database searches to identify an unknown protein. We have incorporated this experiment into an undergraduate general chemistry laboratory for first-year honors students. Many of these students are interested in pursuing undergraduate research projects. This experiment introduces them to a state-of-the-art technique that is commonly used for biochemical research. The experiment is also appropriate for use in a junior- or senior-level undergraduate analytical or biochemistry laboratory, and can be used with relatively little modification to our procedure. In particular, upper-division undergraduates would benefit from a more hands-on experience that would include sample preparation and supervised data acquisition. Throughout this article, we suggest specific modifications to our procedure that would allow this experiment to be used effectively in an upper-level course.

The Experiment

Overview of MALDI–TOF Instrumentation

Before describing the experiment in detail, we provide a brief description of the principles underlying a MALDI–TOF experiment. A schematic illustration of a simple TOF and MALDI setup are shown in Figure 1. In MALDI, the sample is mixed with a matrix, and a droplet of the mixture is spotted on a plate, and allowed to dry. The plate is hit with a very short pulse of light from a laser. Matrix molecules absorb energy from the incident light; this energy is used to volatilize the sample and transfer protons from the matrix to molecules in the sample, instantaneously forming a group of ions. This ionization method favors the production of ions having a single charge.

The ions formed are accelerated into the source region of a TOF mass spectrometer. At intervals, a high voltage pulse is used to send ions into the flight tube (a field-free region), where they separate according to kinetic energies. The kinetic energy (KE) associated with a given ion can be expressed in two ways: with respect to the potential applied or with respect to the resulting velocity of the ions. These relation-

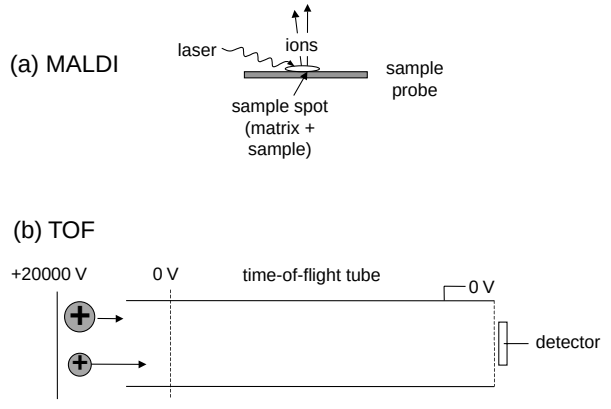


Figure 1. Schematic illustration of (a) MALDI and (b) TOF mass spectrometer. The potentials shown are provided solely as an example.

ships are given by the following equation

$$\mathrm{KE} = zeV = \tfrac{1}{2}mv^2 \quad (1)$$

where z is the charge state of the ion, e is the charge on an electron (1.6022×10^{-19} C), V is the potential drop the ions experience prior to entering the flight tube, m is the mass of the ion, and v is the velocity of the ion. The ion velocity can be expressed as the ratio of the flight tube length, L, and the flight time, t_f: $v = L/t_f$. Combining this relationship with eq 1, one obtains the following equation

$$t_f = L\sqrt{\frac{m}{2zeV}} \quad (2)$$

After a delay period (allowing ions to reach the detector), the next pulse of ions is introduced into the flight tube. A mass spectrum is generated from the sum of spectra recorded for several pulses of ions.

The resolving power (as given by $m/\Delta m$ or $t/\Delta t$) is principally dependent upon the spatial and kinetic energy distribution of ions at the initial pulse. Ideally, when a pulse is applied, all ions would start from a single plane in space that is perpendicular to the line of flight and have the same kinetic energy. Most commercial instruments employ delayed extraction, ion mirroring in a reflectron, and other techniques in order to control the ions' kinetic energy and improve resolving power. We do not discuss these methods here, as our goal is to present a basic overview for our students; however, these topics are appropriate for study by upper-division students.

Preparatory Laboratory Work

In order to demonstrate the instrument operation to students, advance preparation and analysis of samples was required. Briefly, peptide digests were generated *(7)* by incubating solutions containing 2 mM trypsin (TPCK-treated; Sigma, sequencing grade) and 50 mM protein in 100 mM ammonium bicarbonate (JT Baker) for four hours in a 37 °C water bath. All proteins used for digestion were obtained from Sigma and used without further purification. After four hours, the reaction was quenched by adding 2% trifluoroacetic acid (TFA) to comprise 0.1% of the total volume. The matrix was prepared as a 10 mg/mL solution of α-cyano-4-hydroxycinnamic acid (Sigma) in 2:1 water:acetonitrile containing 0.1% TFA by volume. A 1 μL aliquot of the digest solution was combined with 9 μL of the matrix solution. Of this mixture, 1 μL was spotted onto a MALDI sample probe and allowed to dry. The MALDI–TOF used to record data was a Bruker Biflex III, equipped with a reflectron geometry TOF with pulsed ion extraction for high resolution measurements (>10,000 FWHM, ref *8*). Figure 2 shows an example dataset recorded for one of the unknown protein digests (in this case, human hemoglobin).

Hazards

In general, contact with potentially hazardous chemicals in this experiment will be limited to the instructor who performs the preparation of solutions. Protective gloves should be worn for chemical handling. Use of a fume hood and protective eyewear are recommended when handling the solvents trifluoroacetic acid (corrosive) and acetonitrile (an irritant). Hazards may vary depending on the proteins selected for the experiment; in general, care should be taken to avoid inhalation or skin absorption.

Laboratory Instruction

In-class instruction was divided into two parts: a 45-minute classroom lecture, and a demonstration of the experimental procedure in the Indiana University Mass Spectrometry facility. The lecture component included information on proteins, protein digestion, and an introduction to the principles of MALDI–TOF. Students learned about the mechanism of ion formation by MALDI, and ion separation by flight times in the mass spectrometer. Basic equations relating ion kinetic energy, mass, and velocity were reviewed, providing students with a simplified physical understanding of the separation.

In the laboratory, the steps involved in preparing the sample and recording data were demonstrated and explained. Particular emphasis was placed on the sensitivity and speed of the technique. At this stage, most undergraduates have only been introduced to macroscale experiments, and have not been exposed to state-of-the-art methods that are capable of handling the quantity of sample (pmol) required for such an analysis, the methods used to prepare such a sample, or the time required for automated data collection and analysis.

In our course, the demonstration format was chosen because of the size of the laboratory sections, level of students (honors general chemistry), and cost of materials and equipment. Students in an upper-division laboratory course could perform this experiment in small groups over two laboratory periods. The first session would include a brief lecture and allow time for preparation of tryptic digest samples. The digestion protocol could be extended to include other standard preparative steps, such as reduction and alkylation of disulfide bonds (a common step in analysis of large, complex proteins). In the second session, students would be supervised in setting up and performing the MALDI–TOF analysis. An advanced class also permits more freedom in selecting the focus of the experiment; here, we have focused solely on the identification of an unknown sample. An attractive focus for

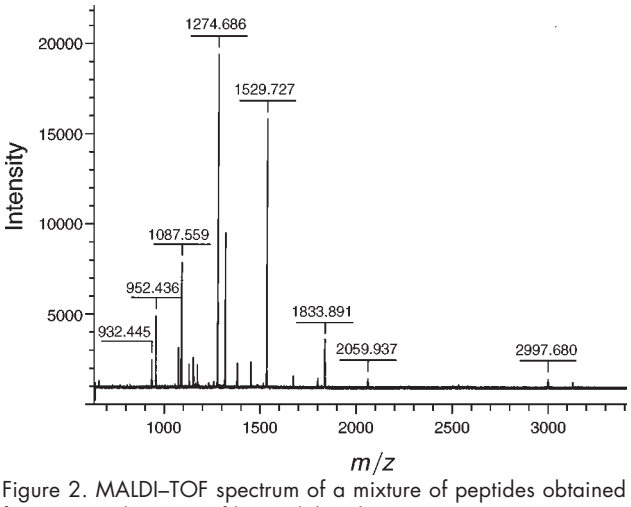


Figure 2. MALDI–TOF spectrum of a mixture of peptides obtained from tryptic digestion of hemoglobin (human).

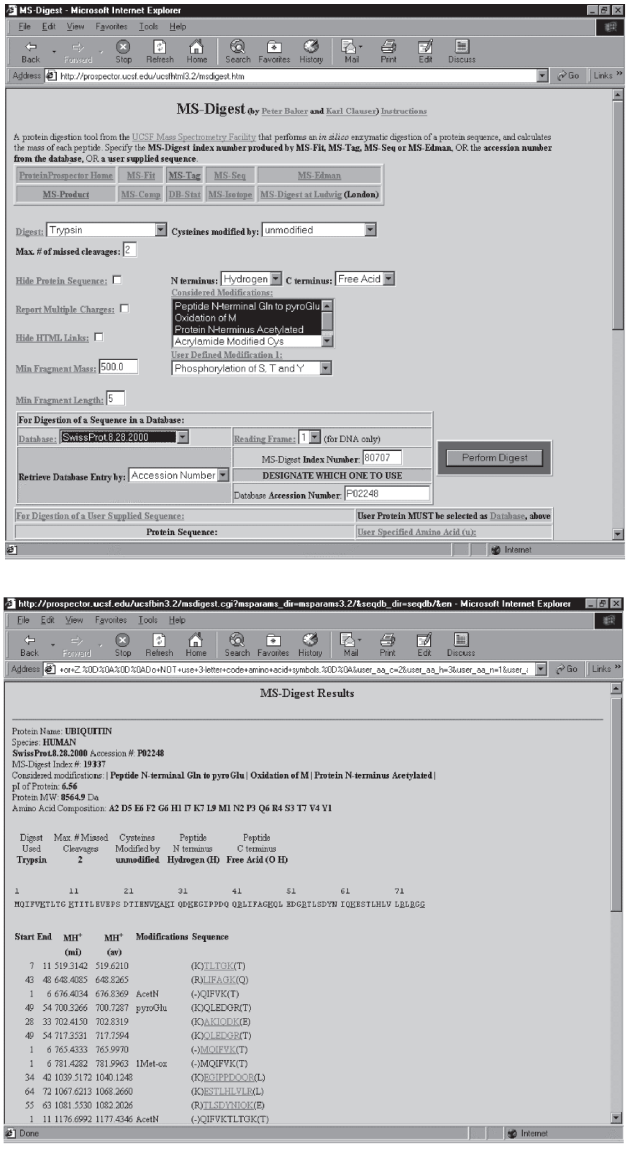
a biochemistry laboratory might be the identification of mutation or posttranslational modification sites by comparing multiple spectra and using database search methods.

Student Assignment: Database Searching

At the end of the in-laboratory demonstration, students were provided with five spectra recorded for tryptic digests: one labeled with the corresponding protein (ubiquitin), and four spectra of unknown protein digests. Detailed instructions on how to use the MS-Digest and MS-Fit tools available on the ProteinProspector Web site *(5)* were provided. For the ubiquitin spectrum, students were asked to identify the sequences associated with the labeled peaks and determine the protein sequence coverage provided by the identified ions. The MS-Digest input form and results page are shown in Figure 3. This exercise underscores the relationship between mass spectra and protein sequences obtained from tryptic digestion.

To identify the unknown proteins, students were provided a list of key parameters necessary to perform the searches on MS-Fit (Figure 4). They were encouraged to vary these parameters to gain a feeling for the robustness of their protein identifications. Once the students obtained a list of protein matches to their spectra, they were asked to examine the database entry for one of the identified proteins, noting information such as protein sequence, origin, and function.

In addition to database searching, the students were required to complete a short problem set, which was based on topics covered in laboratory handouts and lecture. At the honors freshman level, we placed particular emphasis on understanding the physics underlying the mass spectrometric measurement. Students are typically exposed to the equations



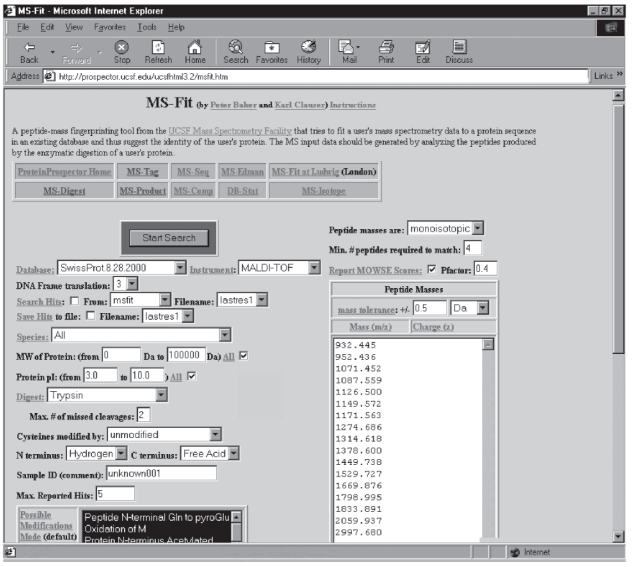


Figure 3. Input form (top) and results page (bottom) from the MS-Digest software obtained for ubiquitin (Swiss-Prot accession number P02248, *http://www.expasy.ch,* ref *6*).

Figure 4. Input form (top) and results summary (bottom) from the MS-Fit search tool (ProteinProspector). The results were obtained using data from the spectrum shown in Figure 1.

that describe ion motion in an electric field in introductory physics courses—in this experiment, we intended to reinforce this understanding and demonstrate the direct application to mass spectrometry.

Summary

The classroom instruction for this experiment provided students with an overview of the operation of a MALDI–TOF instrument and a routine biological application. Students gained an appreciation for the quantity and types of information available for protein characterization on the Internet by performing database searches to identify unknown proteins from mass spectral data.

Supplemental Material

Additional material available on *JCE Online* includes student handouts and instructor's notes. The instructor's notes contain a detailed materials and equipment list and solutions to assigned problems.

Literature Cited

1. Yates, J. R., III *J. Mass Spectrom.* **1998**, *33,* 1.
2. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246,* 64.
3. Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60,* 2299–2301.
4. Muddiman, D. C.; Bakhtiar, R.; Hofstadler, S. A.; Smith, R. D. *J. Chem. Educ.* **1997**, *74,* 1288.
5. ProteinProspector. *http://prospector.ucsf.edu* (accessed Sept 2002).
6. For example, see: MASCOT. *http://www.matrixscience.com* (accessed Sept 2002); MOWSE, *http://www.hgmp.mrc.ac.uk/Bioinformatics/Webapp/mowse/* (accessed Sept 2002); and Peptide Search at EMBL *http://www.mann.embl-heidelberg.de/Services/PeptideSearch/PeptideSearchIntro.html* (accessed Sept 2002). The ExPASy Molecular Biology Server, *http://www.expasy.ch* (accessed Sept 2002), also provides a compilation of complementary protein analysis tools.
7. Houston, C. T.; Reilly, J. P. *Anal. Chem.* **1999**, *71,* 3397.
8. Bruker Daltonik GmbH. *http://www.bruker-daltonik.de/products.html* (accessed Sept 2002).